Human sequence of the non-coding RNA gene (including the putative promoter)

```
   1  CTTAGAGTTT CGTGGCTTCA GGGTGGGAGT AGTTGGAGCA TTGGGGATGT
  51  TTTTCTTACC GACAAGCACA GTCAGGTTGA AGACCTAACC AGGGCCAGAA
 101  GTAGCTTTGC ACTTTTCTAA ACTAGGCTCC TTCAACAAGG CTTGCTGCAG
 151  ATACTACTGA CCAGACAAGC TGTTGACCAG GCACCTCCCC TCCCGCCCAA
 201  ACCTTTCCCC CATGTGGTCG TTAGAGACAG AGCGACAGAG CAGTTGAGAG
 251  GACACTCCCG TTTTCGGTGC CATCAGTGCC CCGTCTACAG CTCCCCCAGC
 301  TCCCCCCACC TCCCCCACTC CCAACCACGT TGGGACAGGG AGGTGTGAGG
 351  CAGGAGAGAC AGTTGGATTC TTTAGAGAAG ATGGATATGA CCAGTGGCTA
 401  TGGCCTGTGC GATCCCACCC GTGGTGGCTC AAGTCTGGCC CCACACCAGC
 451  CCCAATCCAA AACTGGCAAG GACGCTTCAC AGGACAGGAA AGTGGCACCT
 501  GTCTGCTCCA GCTCTGGCAT GGCTAGGAGG GGGGAGTCCC TTGAACTACT
 551  GGGTGTAGAC TGGCCTGAAC CACAGGAGAG GATGGCCCAG GGTGAGGTGG
 601  CATGGTCCAT TCTCAAGGGA CGTCCTCCAA CGGGTGGCGC TAGAGGCCAT
 651  GGAGGCAGTA GGACAAGGTG CAGGCAGGCT GGCCTGGGGT CAGGCCGGGC
 701  AGAGCACAGC GGGGTGAGAG GGATTCCTAA TCACTCAGAG CAGTCTGTGA
 751  CTTAGTGGAC AGGGGAGGGG GCAAAGGGGG AGGAGAAGAA AATGTTCTTC
 801  CAGTTACTTT CCAATTCTCC TTTAGGGACA GCTTAGAATT ATTTGCACTA
 851  TTGAGTCTTC ATGTTCCCAC TTCAAAACAA ACAGATGCTC TGAGAGCAAA
 901  CTGGCTTGAA TTGGTGACAT TTAGTCCCTC AAGCCACCAG ATGTGACAGT
 951  GTTGAGAACT ACCTGGATTT GTATATATAC CTGCGCTTGT TTTAAAGTGG
1001  GCTCAGCACA TAGGGTTCCC ACGAAGCTCC GAAACTCTAA GTGTTTGCTG
1051  CAATTTTATA AGGACTTCCT GATTGGTTTC TCTTCTCCCC TTCCATTTCT
1101  GCCTTTTGTT CATTTCATCC TTTCACTTCT TTCCCTTCCT CCGTCCTCCT
1151  CCTTCCTAGT TCATCCCTTC TCTTCCAGGC AGCCGCGGTG CCCAACCACA
1201  CTTGTCGGCT CCAGTCCCCA GAACTCTGCC TGCCCTTTGT CCTCCTGCTG
1251  CCAGTACCAG CCCCACCCTG TTTTGAGCCC TGAGGAGGCC TTGGGCTCTG
1301  CTGAGTCCAA CCTGGCCTGT CTGTGAAGAG CAAGAGAGCA GCAAGGTCTT
1351  GCTCTCCTAG GTAGCCCCCT CTTCCCTGGT AAGAAAAAGC AAAAGGCATT
1401  TCCCACCCTG AACAACGAGC CTTTTCACCC TTCTACTCTA GAGAAGTGGA
1451  CTGGAGGAGC TGGGCCCGAT TTGGTAGTTG AGGAAAGCAC AGAGGCCTCC
1501  TGTGGCCTGC CAGTCATCGA GTGGCCCAAC AGGGGCTCCA TGCCAGCCGA
1551  CCTTGACCTC ACTCAGAAGT CCAGAGTCTA GCGTAGTGCA GCAGGGCAGT
1601  AGCGGTACCA ATGCAGAACT CCCAAGACCC GAGCTGGGAC CAGTACCTGG
1651  GTCCCCAGCC CTTCCTCTGC TCCCCCTTTT CCCTCGGAGT TCTTCTTGAA
```

Fig. 1

```
1701  TGGCAATGTT  TTGCTTTTGC  TCGATGCAGA  CAGGGGGCCA  GAACACCACA

1751  CATTTCACTG  TCTGTCTGGT  CCATAGCTGT  GGTGTAGGGG  CTTAGAGGCA

1801  TGGGCTTGCT  GTGGGTTTTT  AATTGATCAG  TTTTCATGTG  GGATCCCATC

1851  TTTTTAACCT  CTGTTCAGGA  AGTCCTTATC  TAGCTGCATA  TCTTCATCAT

1901  ATTGGTATAT  CCTTTTCTGT  GTTTACAGAG  ATGTCTCTTA  TATCTAAATC

1951  TGTCCAACTG  AGAAGTACCT  TATCAAAGTA  GCAAATGAGA  CAGCAGTCTT

2001  ATGCTTCCAG  AAACACCCAC  AGGCATGTCC  CATGTGAGCT  GCTGCCATGA

2051  ACTGTCAAGT  GTGTGTTGTC  TTGTGTATTT  CAGTTATTGT  CCCTGGCTTC

2101  CTTACTATGG  TGTAATCATG  AAGGAGTGAA  ACATCATAGA  AACTGTCTAG

2151  CACTTCCTTG  CCAGTCTTTA  GTGATCAGGA  ACCATAGTTG  ACAGTTCCAA

2201  TCAGTAGCTT  AAGAAAAAC  CGTGTTTGTC  TCTTCTGGAA  TGGTTAGAAG

2251  TGAGGGAGTT  TGCCCCGTTC  TGTTTGTAGA  GTCTCATAGT  TGGACTTTCT

2301  AGCATATATG  TGTCCATTTC  CTTATGCTGT  AAAGCAAGT  CCTGCAACCA

2351  AACTCCCATC  AGCCCAATCC  CTGATCCCTG  ATCCCTTCCA  CCTGCTCTGC

2401  TGATGACCCC  CCCAGCTTCA  CTTCTGACTC  TTCCCCAGGA  AGGGAAGGGG

2451  GGTCAGAAGA  GAGGGTGAGT  CCTCCAGAAC  TCTTCCTCCA  AGGACAGAAG

2501  GCTCCTGCCC  CCATAGTGGC  CTCGAACTCC  TGGCACTACC  AAAGGACACT

2551  TATCCACGAG  AGCGCAGCAT  CCGACCAGGT  TGTCACTGAG  AAGATGTTTA

2601  TTTTGGTCAG  TTGGGTTTTT  ATGTATTATA  CTTAGTCAAA  TGTAATGTGG

2651  CTTCTGGAAT  CATTGTCCAG  AGCTGCTTCC  CCGTCACCTG  GGCGTCATCT

2701  GGTCCTGGTA  AGAGGAGTGC  GTGGCCCACC  AGGCCCCCCT  GTCACCCATG

2751  ACAGTTCATT  CAGGGCCGAT  GGGGCAGTCG  TGGTTGGGAA  CACAGCATTT

2801  CAAGCGTCAC  TTTATTTCAT  TCGGGCCCCA  CCTGCAGCTC  CCTCAAAGAG

2851  GCAGTTGCCC  AGCCTCTTTC  CCTTCCAGTT  TATTCCAGAG  CTGCCAGTGG

2901  GGCCTGAGGC  TCCTTAGGGT  TTTCTCTCTA  TTTCCCCCTT  TCTTCCTCAT

2951  TCCCTCGTCT  TTCCCAAAGG  CATCACGAGT  CAGTCGCCTT  TCAGCAGGCA

3001  GCCTTGGCGG  TTTATCGCCC  TGGCAGGCAG  GGGCCCTGCA  GCTCTCATGC

3051  TGCCCCTGCC  TTGGGGTCAG  GTTGACAGGA  GGTTGGAGGG  AAAGCCTTAA

3101  GCTGCAGGAT  TCTCACCAGC  TGTGTCCGGC  CCAGTTTTGG  GGTCTGACCT

3151  CAATTTCAAT  TTTGTCTGTA  CTTGAACATT  ATGAAGATGG  GGGCCTCTTT

3201  CAGTGAATTT  GTGAACAGCA  GAATTGACCG  ACAGCTTTCC  AGTACCCATG

3251  GGGCTAGGTC  ATTAAGGCCA  CATCCACAGT  CTCCCCCACC  CTTGTTCCAG

3301  TTGTTAGTTA  CTACCTCCTC  TCCTGACAAT  ACTGTATGTC  GTCGAGCTCC

3351  CCCCAGGTCT  ACCCCTCCCG  GCCCTGCCTG  CTGGTGGGCT  TGTCATAGCC

3401  AGTGGGATTG  CCGGTCTTGA  CAGCTCAGTG  AGCTGGAGAT  ACTTGGTCAC
```

Fig. 1 (cont'd 1)

```
3451  AGCCAGGCGC TAGCACAGCT CCCTTCTGTT GATGCTGTAT TCCCATATCA
3501  AAAGGCACAG GGGACACCCA GAAACGCCAC ATCCCCCAAT CCATCAGTGC
3551  CAAACTAGCC AACGGCCCCA GCTTCTCAGC TCGCTGGATG GCGGAAGCTG
3601  CTACTCGTGA GCGCCAGTGC GGGTGCAGAC AATCTTCTGT TGGGTGGCAT
3651  CATTCCAGGC CCGAAGCATG AACAGTGCAC CTGGGACAGG GAGCAGCCCC
3701  AAATTGTCAC CTGCTTCTCT GCCCAGCTTT TCATTGCTGT GACAGTGATG
3751  GCGAAAGAGG GTAATAACCA GACACAAACT GCCAAGTTGG GTGGAGAAAG
3801  GAGTTTCTTT AGCTGACAGA ATCTCTGAAT TTTAAATCAC TTAGTAAGCG
3851  GCTCAAGCCC AGGAGGGAGC AGAGGGATAC GAGCGGAGTC CCCTGCGCGG
3901  GACCATCTGG AATTGGTTTA GCCCAAGTGG AGCCTGACAG CCAGAACTCT
3951  GTGTCCCCCG TCTAACCACA GCTCCTTTTC CAGAGCATTC CAGTCAGGCT
4001  CTCTGGGCTG ACTGGGCCAG GGGAGGTTAC AGGTACCAGT TCTTTAAGAA
4051  GATCTTTGGG CATATACATT TTTAGCCTGT GTCATTGCCC CAAATGGATT
4101  CCTGTTTCAA GTTCACACCT GCAGATTCTA GGACCTGTGT CCTAGACTTC
4151  AGGGAGTCAG CTGTTTCTAG AGTTCCTACC ATGGAGTGGG TCTGGAGGAC
4201  CTGCCCGGTG GGGGGGCAGA GCCCTGCTCC CTCCGGGTCT TCCTACTCTT
4251  CTCTCTGCTC TGACGGGATT TGTTGATTCT CTCCATTTTG GTGTCTTTCT
4301  CTTTTAGATA TTGTATCAAT CTTTAGAAAA GGCATAGTCT ACTTGTTATA
4351  AATCGTTAGG ATACTGCCTC CCCCAGGGTC TAAAATTACA TATTAGAGGG
4401  GAAAAGCTGA ACACTGAAGT CAGTTCTCAA CAATTTAGAA GGAAAACCTA
4451  GAAACATTT GGCAGAAAAT TACATTTCGA TGTTTTTGAA TGAATACAAG
4501  CAAGCTTTTA CAACAGTGCT GATCTAAAAA TACTTAGCAC TTGGCCTGAG
4551  ATGCCTGGTG AGCATTACAG GCAAGGGGAA TCTGGAGGTA GCCGACCTGA
4601  GGACATGGCT TCTGAACCTG TCTTTTGGGA GTGGTATGGA AGGTGGAGCG
4651  TTCACCAGTG ACCTGGAAGG CCCAGCACCA CCCTCCTTCC CACTCTTCTC
4701  ATCTTGACAG AGCCTGCCCC AGCGCTGACG TGTCAGGAAA ACACCCAGGG
4751  AACTAGGAAG GCACTTCTGC CTGAGGGCA GCCTGCCTTG CCCACTCCTG
4801  CTCTGCTCGC CTCGGATCAG CTGAGCCTTC TGAGCTGGCC TCTCACTGCC
4851  TCCCCAAGGC CCCCTGCCTG CCCTGTCAGG AGGCAGAAGG AAGCAGGTGT
4901  GAGGGCAGTG CAAGGAGGGA GCACAACCCC CAGCTCCCGC TCCGGGCTCC
4951  GACTTGTGCA CAGGCAGAGC CCAGACCCTG GAGGAAATCC TACCTTTGAA
5001  TTCAAGAACA TTTGGGGAAT TTGGAAATCT CTTTGCCCCC AAACCCCCAT
5051  TCTGTCCTAC CTTTAATCAG GTCCTGCTCA GCAGTGAGAG CAGATGAGGT
5101  GAAAAGGCCA AGAGGTTTGG CTCCTGCCCA CTGATAGCCC CTCTCCCCGC
5151  AGTGTTTGTG TGTCAAGTGG CAAAGCTGTT CTTCCTGGTG ACCCTGATTA
5201  TATCCAGTAA CACATAGACT GTGCGCATAG GCCTGCTTTG TCTCCTCTAT
```

Fig. 1 (cont'd 2)

```
5251  CCTGGGCTTT TGTTTTGCTT TTTAGTTTTG CTTTTAGTTT TTCTGTCCCT
5301  TTTATTTAAC GCACCGACTA GACACACAAA GCAGTTGAAT TTTTATATAT
5351  ATATCTGTAT ATTGCACAAT TATAAACTCA TTTTGCTTGT GGCTCCACAC
5401  ACACAAAAAA AGACCTGTTA AAATTATACC TGTTGCTTAA TTACAATATT
5451  TCTGATAACC ATAGCATAGG ACAAGGGAAA ATAAAAAAAG AAAAAAAAGA
5501  AAAAAAAACG ACAAATCTGT CTGCTGGTCA CTTCTTCTGT CCAAGCAGAT
5551  TCGTGGTCTT TTCCTCGCTT CTTTCAAGGG CTTTCCTGTG CCAGGTGAAG
5601  GAGGCTCCAG GCAGCACCCA GGTTTTGCAC TCTTGTTTCT CCCGTGCTTG
5651  TGAAAGAGGT CCCAAGGTTC TGGGTGCAGG AGCGGCTCCCT TGACCTGCTG
5701  AAGTCCGGAA CGTAGTCGGC ACAGCCTGGT CGCCTTCCAC CTCTGGGAGC
5751  TGGAGTCCAC TGGGGTGGCC TGACTCCCCC AGTCCCCTTC CCGTGACCTG
5801  GTCAGGGTGA GCCCATGTGG AGTCAGCCTC GCAGGCCTCC CTGCCAGTAG
5851  GGTCCGAGTG TGTTTCATCC TTCCCACTCT GTCGAGCCTG GGGGCTGGAG
5901  CGGAGACGGG AGGCCTGGCC TGTCTCGGAA CCTGTGAGCT GCACCAGGTA
5951  GAACGCCAGG GACCCCAGAA TCATGTGCGT CAGTCCAAGG GGTCCCCTCC
6001  AGGAGTAGTG AAGACTCCAG AAATGTCCCT TTCTTCTCCC CCATCCTACG
6051  AGTAATTGCA TTTGCTTTTG TAATTCTTAA TGAGCAATAT CTGCTAGAGA
6101  GTTTAGCTGT AACAGTTCTT TTTGATCATC TTTTTTTAAT AATTAGAAAC
6151  ACCAAAAAAA TCCAGAAACT TGTTCTTCCA AAGCAGAGAG CATTATAATC
6201  ACCAGGGCCA AAAGCTTCCC TCCCTGCTGT CATTGCTTCT TCTGAGGCCT
6251  GAATCCAAAA GAAAAACAGC CATAGGCCCT TTCAGTGGCC GGGCTACCCG
6301  TGAGCCCTTC GGAGGACCAG GGCTGGGGCA GCCTCTGGGC CCACATCCGG
6351  GGCCAGCTCC GGCGTGTGTT CAGTGTTAGC AGTGGGTCAT GATGCTCTTT
6401  CCCACCCAGC CTGGGATAGG GGCAGAGGAG GCGAGGAGGC CGTTGCCGCT
6451  GATGTTTGGC CGTGAACAGG TGGGTGTCTG CGTGCGTCCA CGTGCGTGTT
6501  TTCTGACTGA CATGAAATCG ACGCCCGAGT TAGCCTCACC CGGTGACCTC
6551  TAGCCCTGCC CGGATGGAGC GGGGCCCACC CGGTTCAGTG TTTCTGGGGA
6601  GCTGGACAGT GGAGTGCAAA AGGCTTGCAG AACTTGAAGC CTGCTCCTTC
6651  CCTTGCTACC ACGGCCTCCT TTCCGTTTGA TTTGTCACTG CTTCAATCAA
6701  TAACAGCCGC TCCAGAGTCA GTAGTCAATG AATATATGAC CAAATATCAC
6751  CAGGACTGTT ACTCAATGTG TGCCGAGCCC TTGCCCATGC TGGGCTCCCG
6801  TGTATCTGGA CACTGTAACG TGTGCTGTGT TTGCTCCCCT TCCCCTTCCT
6851  TCTTTGCCCT TTACTTGTCT TTCTGGGGTT TTTCTGTTTG GGTTTGGTTT
6901  GGTTTTTATT TCTCCTTTTG TGTTCCAAAC ATGAGGTTCT CTCTACTGGT
6951  CCTCTTAACT GTGGTGTTGA GGCTTATATT TGTGTAATTT TTGGTGGGTG
```

Fig. 1 (cont'd 3)

```
7001    AAAGGAATTT TGCTAAGTAA ATCTCTTCTG TGTTTGAACT GAAGTCTGTA

7051    TTGTAACTAT GTTTAAAGTA ATTGTTCCAG AGACAAATAT TTCTAGACAC

7101    TTTTTCTTTA CAAACAAAAG CATTCGGAGG GAGGGGGATG GTGACTGAGA

7151    TGAGAGGGGA GAGCTGAACA GATGACCCCT GCCCAGATCA GCCAGAAGCC

7201    ACCCAAAGCA GTGGAGCCCA GGAGTCCCAC TCCAAGCCAG CAAGCCGAAT

7251    AGCTGATGTG TTGCCACTTT CCAAGTCACT GCAAAACCAG GTTTTGTTCC

7301    GCCCAGTGGA TTCTTGTTTT GCTTCCCCTC CCCCCGAGAT TATTACCACC

7351    ATCCCGTGCT TTTAAGGAAA GGCAAGATTG ATGTTTCCTT GAGGGGAGCC

7401    AGGAGGGGAT GTGTGTGTGC AGAGCTGAAG AGCTGGGGAG AATGGGGCTG

7451    GGCCCACCCA AGCAGGAGGC TGGGACGCTC TGCTGTGGGC ACAGGTCAGG

7501    CTAATGTTGG CAGATGCAGC TCTTCCTGGA CAGGCCAGGT GGTGGGCATT

7551    CTCTCTCCAA GGTGTGCCCC GTGGGCATTA CTGTTTAAGA CACTTCCGTC

7601    ACATCCCACC CCATCCTCCA GGGCTCAACA CTGTGACATC TCTATTCCCC

7651    ACCCTCCCCT TCCCAGGGCA ATAAAATGAC CATGGAGGGG GCTTGCACTC

7701    TCTTGGCTGT CACCCGATCG CCAGCAAAAC TTAGATGTGA GAAAACCCCT

7751    TCCCATTCCA TGGCGAAAAC ATCTCCTTAG AAAGCCATT ACCCTCATTA

7801    GGCATGGTTT TGGGCTCCCA AAACACCTGA CAGCCCCTCC CTCCTCTGAG

7851    AGGCGGAGAG TGCTGACTGT AGTGACCATT GCATGCCGGG TGCAGCATCT

7901    GGAAGAGCTA GGCAGGGTGT CTGCCCCCTC CTGAGTTGAA GTCATGCTCC

7951    CCTGTGCCAG CCCAGAGGCC GAGAGCTATG GACAGCATTG CCAGTAACAC

8001    AGGCCACCCT GTGCAGAAGG GAGCTGGCTC CAGCCTGGAA ACCTGTCTGA

8051    GGTTGGGAGA GGTGCACTTG GGGCACAGGG AGAGGCCGGG ACACACTTAG

8101    CTGGAGATGT CTCTAAAGC CCTGTATCGT ATTCACCTTC AGTTTTTGTG

8151    TTTTGGGACA ATTACTTTAG AAAATAAGTA GGTCGTTTTA AAAACAAAAA

8201    TTATTGATTG CTTTTTTGTA GTGTTCAGAA AAAGGTTCT TTGTGTATAG

8251    CCAAATGACT GAAAGCACTG ATATATTTAA AAACAAAGG CAATTTATTA

8301    AGGAAATTTG TACCATTTCA GTAAACCTGT CTGAATGTAC CTGTATACGT

8351    TTCAAAAACA CCCCCCCCCC ACTGAATCCC TGTAACCTAT TTATTATATA

8401    AAGAGTTTGC CTTATAAATT TA
```

Fig. 1 (cont'd 4)

Murine sequence of the non-coding RNA gene (including the putative promoter)

```
   1  CTTAGAGTTT CGTGGCTTCG GGGTGGGAGT AGTTGGAGCA TTGGGATGTT
  51  TTTCTTACCG ACAAGCACAG TCAGGTTGAA GACCTAACCA GGGCCAGAAG
 101  TAGCTTTGCA CTTTTCTAAA CTAGGCTCCT TCAACAAGGC TTGCTGCAGA
 151  TACTACTGAC CAGACAAGCT GTTGACCAGG CACTCCCCCC AACAATATCC
 201  TCCCTCTTCC CCCCCCCCAC CCCCGCCCCG TGTGCTCGTT AGGGCAATTG
 251  AAAGGACACT CCCATTTTTG GTGCCATTGA TGCCCTGTCC ATAATAGCTT
 301  CCCTGACTTT TACACCACCC CAACTCCCAA TCTGAAGGAC TGGGAGGTGT
 351  GATGCAGGAG AAACTATGGG ACTCTTGGGA GAAGACTATG GAGTTGGCCA
 401  GTGATTAAGG CCCACTAATT CCAACTGTGG TAGCACAGAT CTGGCTCCAC
 451  ATCAACCCAA TCCAAAACTG ACAAGGATAT TTTGCAAAAA AAGAAAGTGG
 501  CACCTGTCTG ATCCAGCTCT GACATGGCTA GAGGTGAGTC CTAAACTGAT
 551  GGCTTATAAA CTAGCCTGAG CCACAGAAGA GTATGGCCCA GAGTGAAGTG
 601  TCATCATCTG TTCACAAGGC ATGCTCCCCT AGAAGATAAT GCTAAAGAGG
 651  TGCCATGGAG GCAGCAGGAC AAAGTACAGG CAGGCTAGGT GGAGTCAAGC
 701  CAGGCCTAGT GCCACAGAAC AAGAGAGCAG TCTGACTAGT AATTAAGAGG
 751  GAAGAAAGGA AAATATTCTT CCAATTACTT TCCAGTTCTC CTTTAGGGAC
 801  AGCTTAGAAT TATTTGCACT ATTGAGTCTT CATGTTCCCA CTTCAAAACA
 851  AACAGATGCT CTGAAAGCAA ACTGGCTTGA AATGGTGACA CTGTCCCACA
 901  AGCCACCAGA CATGGCAGTG TTCAGAACTA CCTGTATCTG TATATACCTG
 951  CGCTTGTTTT AAAGTGGGCT CAGCACATAG GATTCCCAAG AAGCTCCGAA
1001  ACTCTAAGTG TTTGCTGCAA TTTTATAAGG ACTTCCTGAT TGCTTTCTCT
1051  CTCGTCCTTC CATTTCTTCC TTCCTTCCAT TTCATGCTTT CATTTCTTCC
1101  CCTAGCTTCT AGTTGTTTCT TCTGTTCCAG GCAGCTGCAG TGCTGAACCA
1151  CATGGTTACC TAACAGCAGT CAGCTGCAGC CCTAGGATTC TTCCTGCCCT
1201  TTAACTTCCC ATTGCCAGTG CCAGGTATCA TATTTAACCT TGAGCAAGAG
1251  CTGGGCTCTT TTGAGCCCTC CCTAACCTCT GTGAAGAAGA ACAAGAAGGT
1301  AGGAAGCTCT TGCTCTTGCT AAGAAAAATG TCAAAAGGCT TTCAGACCTT
1351  AAACAATGAG CCTTTTCACC TTTTACTCTA GAAAAGTGGA CTAGAAAATC
1401  TGGGTCACAT TGGGTAGCTG AAGGAGATAC AGAGGCCCCT ATGGCCTGCC
1451  AGAGTCGTTG CATGGCCCAA CAGGGGCTCC ATGCCCACTA CCCTTGACCC
1501  TACTCAGAAA TCTAATGTCA TACTTAGTGT GGGCAGGGGA CCTGTCAGGA
1551  CAGATGCAGA CCTAAGCAGG GAGTGACACC AGGGCCCTTG GCCCTTCTTC
1601  TGACAAACAT ACACATCCCA AGTCTTTTTC TAGTGGAATT CTTAACCTCT
1651  TGCTCACTGG GGACTGGGAA GCATCAGCAC ATCCCATATT TCAAACTCTG
```

Fig. 2

1701   CTCCATAAGT ACAGTGGTGA ATTTTATAGA CTTGACTTTG CTGTGGGGTT

1751   TTAATTGGTC AGTTTTAATT TGGGATCCCA AAGTTTTAAC CTCCATTCAG

1801   GAAGTCCTTA TCTAGCTGCA TATCTTCATC ATATTGGTAT ATCCTTTTCT

1851   GTGTTTACAG AGATGTCTCA TATCTATCGA AATCTGTCTG AGAAGTACCT

1901   TATCAAAGTA GCAAATGAGA CAGCAGTCTT ATGCTTCCAG AAACACCCAC

1951   AGGCACGTCC CATGTGAGCT GCTGCCATGA ACTGTCGAGT GTGTATTGTC

2001   TTGTGTATTT TCGTTAACGT TCCCCAGCTT CCTTCCTGCG GTGTAATCAT

2051   GGAAGAGTGA AACATCATAG AAATCGTCTA GCACTTCCTG GCCAGTCCTT

2101   AGTGATCAGG AACCGTAGTT GACAGTTCCA ATTGATAGCT TAAGATAAAA

2151   CCATGTTTGT CTCTTATGGA ATGGTTAGAA CTAAGTGAGA GATCTTGCCC

2201   CATTCTGTTT GCCGAATCAT AGTTGGACTT TTAGTGTATT TGTATCCATT

2251   TCCTTGTGCT ATAAAAGCAA ACCCTGCAAC CAGCTTTCTG TCAGGCAGTC

2301   CTTTTGCCTG CTCTGCTTTT GATCCTCTTA GTCTTGCTTC TGGTTCCTCC

2351   CTGGAGAGGG AGGAGGGGTC AGAAGAGGAA TTCTGGAGGA TCCAGGATAT

3401   GTCCTTCTGA ACTCCTGCTT CTTCCAGTGA CAAAGGCCC CTACTGCCCC

2451   ACCCCAACCT GCCCCATGCA CTCCTCTAGG ACACCTTTCC ATACTTTTCA

2501   CAACACCTAG CCAGGTTGAC ACCAAGTTGT TTATTGTGGT CTGCTTGGAA

2551   TTTTACCTGT TAGGCTTACT TAGTCCAATC AAATGGACTC CAAGTTGGGT

2601   ATCCCTCATC TTTGGAAGAC AACCTAGGCT GATTAGATAT TTACTTTTGG

2651   GATTGCAGCA CTTTGGGTGC CGTTTTTCTT TTACTTGGGT TTTATCTGCA

2701   GCTCCCTCAC CACCACCACC ACCCCCCACT TACCTGTATG TAGAACTGAT

2751   TTCAAAACTG CAGGTGGTGG TAACTGCAGC TTCTTAGGGT TTTCTTCACT

2801   TCTTGCTTCT TTCCCCATTC CCTCATCCAC AAATAAGGGC ATCACAAGTC

2851   AGTCTCCTTT AAGCAGGCAG CTTTGGTGGG GTTTTTCCCC TGGAAGCCAG

2901   GGACCCTGTC AGGCTGCCTC TGCCTTGTGG TCAGGTTGAC AGGAGGTTGG

2951   AGGGAAAAGC CTTAAGTCAT GGGATTCTCA CCAGCTGTGT CTGGCTCAGA

3001   CCTGGAATGT GACCTTTATT TTGTTGTATT TGAACATTGT AAAGTGTGGG

3051   TGGTACCTTA AACTGAATAT GTGAAGAATC CAGAAACTGA CCAACAGCTT

3101   TCAGATACCT GGGGCTAGGT CACTAAGGTC ACATCCAGTC TTCCCTACCC

3151   TGTTCTAGTT GTTAGCTACT ACCTCTCCCA GATAGATTGC TGTATATCCT

3201   CCAACTATGA TCATCCTGGC CCAAGCTTGC CTGTTCTTGA GTCTGTCTTA

3251   ACCAGTGGAA CTGCTGCCCT TGGTGTGCAG TGAGTTGAGG ACTCTTGGTC

3301   ACAGCCAGGC TCTAGTAGTA CAGCTCCTTT CTGCTGGTGC TGTATTTCCA

3351   TATCAAAAGG CACAGGGGAG ATCTAGAAAT GCCATCTCCC CCAGTCCATC

3401   AGTGCCAAAC AAGCCCATGA TCCCAGCATG GGTACAGACA ACTCTGTTCA

Fig. 2   (cont'd 1)

```
3451  GTGCTATCAC AACAGACTAG AGGCCATGAA CATTGGACGT GGGAACCAGA

3501  GCAACCCGAA TTGCTGCTGC TTTATTCAGC TTTCCGTTGC TCTGACAATG

3551  ATAAACAAG GCAGTAACTT AAAACAGACT GCCAGGTTTG GCAGAGAAAG

3601  GAAATTCCTT AGCTGACAGC ACCTCTGGAT TTTAAATAGG TTGTAATAAG

3651  TGGCTCAAAC CCATCCAGGA AAAAGCAAAA GGGTTAGAAC TGACCAGATG

3701  AGACCAGCCT GATTTCATGC AGCCCAAATG GAGTCCAGCT GTCTGAACTC

3751  TGCAGCACTT CTCTACTACA GTCTCCTAGA GCATTCCAGC CAGGCTCTTC

3801  AGGCTGAGGA GACATCACAG GTGCCAGTTC TTCAAGAAGA CTTTTGTGCA

3851  TCAGTTCATA GCCTATATCT TTGCCCAAGA TTGTAGATTC AGGTTAACAC

3901  TACAGATTCT AGGGCAGATG ACTGAGACTC AGAAAAAAAG CCCCTGTGGA

3951  CTGTGGTATA GCGAAGTACA AAAACTGAAG GGGGCTAGGG CAGATGCCGC

4001  ATGCCTCATG CCAGAGCCAA GCCCTCTGCT CCATCCACAT CCTTTTCTGG

4051  CTCCTTCTTC CTGCTCTCTG CTTCAGTGAA CCAGCCCCAC TCTGAAGAGA

4101  TTTGTTGATT CTCTCCATTT TTATGTCTTT CTCTTTTAGG TACTATATAG

4151  AAAAGGCTTA GTCTAATTGT TATAAATTGC TAGAATACTG CCTCCCCCAG

4201  GGTCTAAAAA TATATGCTAA AGGGGAAAAC TTGAACACTG AAACCAGTTC

4251  TGAACAATTT AGAAGGAAAA CCTTGAAAAC ATTTAACAAA AAATTATATT

4301  TTAATGTTTA TGAATAAGAG GAGGCTTTTG AAAAAATGTT GATCTATAAA

4351  TACTTACTTT AGGCCTGAGG TGTCTAATGA GTGAACTGAG CAATGGGAAC

4401  TCAAGGCTGA AGCCTCCTGC ATCAGAGGAG GTAGAACCAG GAGCCTCTTG

4451  AGATTTGAGG TGTTTTAGCA TTGGAAAGCC ACTCTTTGGG TAGCTGGCCC

4501  CAGAAACTAC TTCTGACCTT GTCATTTGGA ATGGAGGTTA GTGGTCTGCC

4551  AGATGCCAAA GCTGCATGAG ACCAGCTCTT GGTTTATCAA TTTGAACACT

4601  CAGTAACCTA GAAGGCCCAG CACAAAGTGT CTGCTCTCTT CTTAACTGAG

4651  CCTGCCCCAG CACTACTGCA CAAATTAGGG AGGGTCTACT TCCTACAGAG

4701  CATCCCTCCC TGGGCCCCCT CCCATCCTTT GTACTCTACC TACCTGACCT

4751  TCAGGATCTT GGCACATACG AAATGGCTGT GTAGCAAGCA CTTTGGCATG

4801  CCCTCCTAAA CTTACCCCAG AGCCTCTCCC TGCCTCCTTA AGCCAGTCTG

4851  CCTGTCTTCT GGGGAGGTGT TAGAGCCCAT AGAATGGAGA GGAGAAAGAA

4901  AAGAGGAAGA GGCAGGCAGG TAGTAAAAAG GCTCTGGGAG GAAAGACAGC

4951  CTCCTAGGCT TTGCACAAGC AGGACTCAGC CCCTTGTGGG AACTAAGTGC

5001  CATCTTGGAG TTTAAGAACA TTTGGACAAG TTGCAAATGA CCTTTGCTCC

5051  TTGCTCCTCT CACCTTTTAT GGGGCCCTGC TTAGCACTGA AAGCAAATGC

5101  GCTGAAAAGG CAAAGAGGTT TGGCTCCTGC CCACTGATAG TCCTTTCCCT

5151  GCAGTGTTTG TGTGTCAAGT GGCAAAGCTG TTCTTCCTGG TGACTCTGAT

5201  TAGATCCAGT AACTTAAGAG ATTTGTATGC ATAGGTCTGC TTTGACTCTT
```

Fig. 2 (cont'd 2)

```
5251  CTATTCTGGG CTTTTGATTT GTTTTTCAGT TTTGCTTTTA GTTTTCCTAT
5301  TTTTATTTTA TGCACCAACT AGACACACAA AGCAGTTGAA TTTATATATA
5351  TATATATATA TATATATCTG TATATTTCAC AATTATAAAC TCATTTTGCT
5401  TGTGACGCCA CACACACACA AAAAGAAAAA CCTTTTAAAA TTATACCTGT
5451  TGCTTAATTA CAATATTTCT GATAACCATA GAGTAGGACA AGGGAAAAAA
5501  TTTAAAAAAA AAAAAAAAAA AAGAAAAAC ACATCTGTCT GCTGGTCACT
5551  TCTTCAATCC AAGCAGATCT GTGATCTTTC CTCGCGTCTT TCAAAGACTT
5601  CCCTGTGCTA AGTGAAGGAA GCTCCAGGCT GCACCCAGGT TTTGTGCTTT
5651  GTTTCTCCTC TGTTGTGAAA GGGGCCCCAA GATTCTGGGT ACAGGACAGT
5701  TCATTTCAGC ATGGGGTCAG GAGACAAGAG CACTCCCTTT ACATGCTGAC
5751  GTACAGAACT TAGTGGGAAT AGCCTAGTCC CCACCTCTAG GGATGGGGAG
5801  CTAGCATGCA TGGGGGTGAC CCAACTCCCT CCACCTTTCC CTGGCCAGGA
5851  AGAGCCTGTG TACAGTAAGT CTGACAAGCT TTCCCCAGTT AGCAGGGCTC
5901  AGAGCATTTA AAAACCCTCC AAACTTTGCT GAGTCTAGGG ACTAGAGAGA
5951  AGATAGAAGA TTTGGTCTAT CTCCAAGGTG TGTAAGCTGT ACCAGGTAGA
6001  ATGCCAGGGA CCCCAGAACC ACATCCAACA GCCCAATGGG TCTCCTCCAG
6051  AAAGTAGTGA AGACTCCAGA AACATCCCTT TCTCTTCTCC CTGCTCCCAT
6101  GAGTAACTGC ATTTGCTTTT GTAATCCTTA ATGAGCATTA TCTGCTAAAA
6151  AAAAAAATT AGCTGTAACA GTTCTTTTTG CAAAAGGATC ATTCTTAAAT
6201  AATTAAAAAC ACCCCCCCCC CAAAAAAAAG TCCAGAACCT TGTTCTTCCA
6251  AAGCAGAGAG CATTATAATC AGGGCCAAAA TCTGTCCCAC ACCTCTACCC
6301  CATCTCCTCA TGATTGCTGC TTCTAAGGCC AGAATACAGC AAAGATATTT
6351  GTAGGCCCTT TGGGTGACTG GGCTACCCTT GGAGCTCTTG GAAGATGGGC
6401  TGGGGAAGCC TCTGAGACCC TATCCTAGGG CCTTGCTCTA GGGAGTAATC
6451  AGTATTAGTA GAGTGTCACA ACATTATTCC CCAGCCGGCA TGAGATGGGG
6501  GCAGAAGAAG CCAAAGGGTT GTCTCCACTG CTACTTACTT GGCCACTGAC
6551  AGGTAGGTGA CCATGTATGT CCATATGCAT GTTTTATGGC TGATGTGAGA
6601  TCAGCACCCA AGTTAGCTTC ACCTGGTGAC CTCTAACCCT GCCTGGATGG
6651  AGCAGGCCAC CTGGTTCAAT GTTTCTGGGC AGCTGGACAA TGGAGTGCAA
6701  AAGGCTTACA GAACTTGAAG CCTTTTCCTT ACTTTGCTAG CACGGCCTCC
6751  TTTTCCATTT GATTTGTCAC TGCTTCAGTC AATAACAGCC GCTCCAGAGT
6801  CAGTAGTTGA TGAATATATG ACCAAATATC ACCAGGACTG TTACTCAACG
6851  TGTGCCGAGC CCTTTCCTTG TGCTGGGCTC CCTGTGTACC TGGACACTGT
6901  AATGTGTGCT GTGTTTGCTC TCCTTCCTCT TCCTTCCTTG CCCTTTCCTT
6951  GTCTTTCTGG GGTTTTTCTG TTGGGTTTGG TTTGGTTTTA TTTTTCCTTT
```

Fig. 2 (cont'd 3)

```
7001   TGTGTTCCAA ACATGAGGTT TTCTCTACTG GTCCTCTTTA ACTGTGGTGT

7051   TGAGGCTTCT ATTTGTGTAA TTTTTGGTGG GTGAAAGGAA CTTTGCTAAG

7101   TAAATCTCTT CTGTGTTTGA AATGAAGTCT GTATTGTAAC TATGTTTAAA

7151   GTAATTGTTC CAGAGACAAA TGCTTCTAGG TACATTTTCA TTACAAACAA

7201   AGCATTTGAA GGGAGGGAAG TGGTGAATAA GACAAGAGGG GCAATCTGAA

7251   TTGATCCCTG CCCAGATCAG CCAGAAGCTA CCAAAGTTA AGCACTGGTT

7301   TTCCATTCCA AGTCAAGAGA CTGAAGCTGA TGTTTTGCCA TTTTCAAAGT

7351   CAAAGCAAAA CCAGCTTTTC CACCCAATGG ATTCTTTGCT TCTCCTTCCC

7401   AGATTATTAC TACTGCTGTA ATAATCTAGG AGTGCCAGGA GGGAAAGGAG

7451   TATTAACACA GAGCTGTGCT CACTGAGTAT GGAAAGGCTT GGTCTGAGTT

7501   TTCAGGAGGA TGACCCACTG TGGACATGGG GAGAAGACAG AAGATAAATT

7551   AGCCGCTCCC TGCCTAAGAT ACCTCTTAAT AGATAAGTCA AGGCCATGGA

7601   CATTATTGTC TACAAGGCAT GTTTCAAAGA CATGACCAGT CAGGACACTT

7651   CTGTCATACT CCATGTTGCC CCCTAGTACA CAGTACTAAT CTGATATCTC

7701   TGTTCCCGCC ATGCCTGGGG GATAAAATGA TAGCAGAGAC TCCTTTCCTT

7751   CAATGTGATC TAATTCCCAA CAAATCTGG GCCTGAGATA CCACCTGTTT

7801   CTATGGCAAA CATCCTCAGT AAAGTGTTAT TCTCATTGCA GATTGTTCCA

7851   GCCTAATGTA AGAGGAACAG AGCAGTGTTC CCTTGGAGCC TCATGTGGAC

7901   AGTTCTACCT GTAGTGACCA GTTGGCTATA GTAGTTATTA GCTGGAACAA

7951   CCAGACAGGG TACATGCCCC CTCCAAAATC CATGTTGTAC TCCCCTCTGC

8001   CAGCCAGGGG GGGTGAGATC TGTAGAATAG TGCAGCCAGT GACAAGCCAC

8051   CTTGTGTTTG TCACCAGCTC AAAAACTCAT CTAAGGTTGG GAGCAGGCAG

8101   ACAAGGCAGA GAGAAAGATC CAGGACAGAC CTAGCTGGGC TGGAGGGGTC

8151   TTGAAAAGCC CTCTGTCGTA TTCACCTTCA GTTTTTGTGC TTTGGGACAA

8201   TTACTTTAGA AAATAAGTAG GTCGTTTTAA AAACAAAATA TTGATTGCTT

8251   TTTTGTAGTG TTCAAAACAA AAGGTTCTTT GTGTATAGCC AAATGACTGA

8301   AAGCACTGAT ATATTTAAAA ACAAAGGCA ATTTATTAAG GAAATTTGTA

8351   CCATTTCAGT AAACCTGTCT GAATGTACCT GTATACGTTT CAAAAACACA

8401   CCCCACTGAA CCCCTGTAAC CTATTTATTA TATAAAGAGT TTGCCTTATA

8451   AATTTACATA AAAA
```

Fig. 2 (cont'd 4)

09/720215

Human
Mous

```
   1 CTTAGAGTTTCGTGGCTTCAGGGTGGGAGTAGTTGGAGCATTGGGGATGT
   1 ----------------------G----------------------.-------

  51 TTTTCTTACCGACAAGCACAGTCAGGTTGAAGACCTAACCAGGGCCAGAA
  50 --------------------------------------------------

 101 GTAGCTTTGCACTTTTCTAAACTAGGCTCCTTCAACAAGGCTTGCTGCAG
 100 --------------------------------------------------

 151 ATACTACTGACCAGACAAGCTGTTGACCAGGCACCTCCCC..........
 150 ------------------------------------TC----CAACAATATC

 191 ........TCCCGCCCAAACCTTTCCCCCATGTGGTCGTTAGAGACAGA
 200 CTCCCTCTTC---C---CC---CCCG----G----C----A---

 232 GCGACAGAGCAGTTGAGAGGACACTCCCGTTTTCGGTGCCATCAGTGCCC
 241 .....--G---A----A----------A----T--------TGA-----

 282 CGTC...TACAGCTCCCCAGCTCCCCCCACCTCCCCCACTCCCAACCAC
 286 T---CATA-T----T---TGA--.TTTA----A----A--------..T-

 329 GTTGGGACAGGGAGGTGTGAGGCAGGAGAGACAGTTGGATTCTTTAGAGA
 333 TGAA----T----------T--------A--TA-G--C----GG---A

 379 AG...ATGGATATGACCAGTGGCCTATGGCCGGATCCCACCCGTGGT
 383 --ACT-----GT--G-----AT--A---.CACTA--T---A-T-----

 426 GGCTCAAGTCTGGCCCCACACCAGCCCCAATCCAAAACTGGCAAGGACGC
 432 A--A--GA------T-----T--.A--------------A------TAT

 476 TTCACAGGACAGGAAAGTGGCACCTGTCTGCTCCAGCTCTGGCATGGCTA
 481 --TG--AA-A-A-----------------A---------A-------.

 526 GGAGGGGGGAGTCCCTTGAACTACTGG.GTGTAGACTGGCCTGAACCACA
 530 ...A--T------A--.-----GA----CT-A--A----A---------

 575 GGAGAGGATGGCCCAGGGTGAGGTGGCATGGTCCATTCTCAAGGGACG.T
 576 -A----T--------A----A---T--CA--TG--A-----C-T-C-

 624 CCTCCAACGGGTGGCGCTAGAG....GCCATGGAGGCAGTAGGACAAGGT
 626 --C-T-GAA-A-AAT----A--AGGT------------C-------A--

 670 GCAGGCAGGCTGGCCTGGGGTCAGGCCGGGCAGAGCACAGCGGGGTGAGA
 676 A----------A-.G---A----A---A---CT--TG-CA-A-AACA--.

 720 GGGATTCCTAATCACTCAGAGCAGTCTGTGACTTAGTGGACAGGGGAGGG
 724 -------------------------ACTAG--A--------

 770 GGCAAAGGGGGAGGAGAGAAAAAATGTTCTTCCAGTTACTTTCCAATTCTC
 744 ...T---A---A--A-G------A--------A---------G-----

 820 CTTTAGGGACAGCTTAGAATTATTTGCACTATTGAGTCTTCATGTTCCCA
 791 --------------------------------------------------

 870 CTTCAAAACAAACAGATGCTCTGAGAGCAAACTGGCTTGAATTGGTGACA
 841 --------------------A----------A-------A---------

 920 TTTAGTCCCTCAAGCCACCAGATGTGACAGTGTTGAGAACTACCTGGATT
 891 C--..-----A----------CA--G------C---------T--C

 970 TGTATATATACCTGCGCTTGTTTTAAAGTGGGCTCAGCACATAGGGTTCC
 939 ---..-----------------------------------------A---

1020 CACGAAGCTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCC
 987 --A-----------------------------------------------

1070 TGATTGGTTTCTCTTCTCCCCTTCCATTTCTGCCTTTTGTTCATTTCATC
1037 ------C-------CTCGT------------T-----CCT-C--------G

1120 CTTTCACTTCTTTCCCTTCCTCCGTCCTCCTCCTTCCTAGTTCATCCCTT
1087 ------T-----C----AG-...............-T------.G-TT---

1170 CTCTTCCAGGCAGCCGCGGTGCCCAACC.............ACACTTGTC
1122 --G-----------T--A----TG-----ACATGGTTACCTA---GCA---

1207 GGCTCCAGTCCCCAGAACTCTGCCTGCCCTTTGTCCTCCTGCTGCCAGTA
1172 A---G---.---T--G-T---T---------AA-T---CAT-------G

1257 CCAGCCCCACCCTGTTTTGAGCCCTGAGGAGGCCTTGGGCTCTGCTGAGT
1221 -----...GT-T-A-A---..--A--T----C-A-AGC-------TT----C

1307 CCAACCTGGCCTGTCTG.TGAAGAGCAAGAGAGCAGCCAAGGTCTTGCTCT
1267 --TC---AA----C-G--AA-----A------AG-T--G---C--------

1356 CCTAGGTAGCCCCCTCTTCCCTGGTAAGAAAAA..GCAAAAGGCATTTCC
1317 .................---C---------TGT--------.----A

1404 CACCCTGAACAACGAGCCTTTTCACCCTTCTACTCTAGAGAAGTGGACTG
1345 G---T-A-----T----------T---------A---------.---A

1454 GAGGAGCTGGGCCCGATTTGGTAGTTGAGGAAAGCACAGAGGCCTCCTGT
1394 --AA-T-----T-AC---G-----C---A-G-GAT---------.---A-

1504 GGCCTGCC..AGTCATCGAGTGGCCCAACAGGGGCTCCATGCCAGCCGAC
1443 --------AG----G-T-CA----------------------CA-TAC-

1552 CTTGACCTCACTCAGAAGTCCAGAGTCTAGCGTAGTGCAGCAGGGCAGTA
1493 -------CT-----A--T-AT---ATA-T-----....T-------..

1602 GCGGTACCAATGCAGAACTCCCAAGACCCGAGCTGGGACCAGTACCTGGG
1537 ..--G---TG-CAG--CAGATGC----TA---A----GTGAC---A--

1652 TCCCCAGCCCTTCCTCTGCTCCCCCTTTTCCCTCGGAGTTCTTCTTGAAT
1585 C--TTG-------T----ACAAA-A-ACA--ATC-CA---CT--T-CT-G-

1702 GGCAATGTTTTGCTTTTGCTCGATGCAGACAGG...GGGCCAGAACACCA
1635 --A-T-C--AAC--C------AC--GG---T--GAA-CAT---C---T-C

1749 CACATTTCACTGTCTGTCTGGTCCATAGCTGTGGTGTAGGGGCTTAGAGG
1685 --T-------...AAC---C------AG-ACA---GT-AATT---T--A

1799 CATGGGCTTGCTGTGGGTTTTTAATTGATCAGTTTTCATGTGGGATCCCA
1731 -T--ACT----------G--------G----------A-T---------
```

```
1849 TCTTTTTAACCTCTGTTCAGGAAGTCCTTATCTAGCTGCATATCTTCATC
1781 AAG----------CA-----------------------------------

1899 ATATTGGTATATCCTTTTCTGTGTTTACAGAGATGTCTCTTA..TATCTA
1831 -------------------------------------A--TC----G-

1947 AATCTGTCCAACTGAGAAGTACCTTATCAAAGTAGCAAATGAGACAGCAG
1881 --------------------------------------------------

1997 TCTTATGCTTCCAGAAACACCCACAGGCATGTCCCATGTGAGCTGCTGCC
1927 ----------------------------C---------------------

2047 ATGAACTGTCAAGTGTGTGTTGTCTTGTGTATTTCAGTTATTG.TCCCTG
1977 ----------G---------A---------------TC----AC-T----CA

2096 GCTTCCTTACTATGGTGTAATCATGAAGGAGTGAAACATCATAGAAAACTG
2027 ---------C--GC--------G-A---------A-----------TC-

2146 TCTAGCACTTCCTTGCCAGTCTTTAGTGATCAGGAACCATAGTTGACAGT
2077 --------------G------C------------G-----------

2196 TCCAATCAGTAGCTTAAGAAAAAAACCGTGTTTGTCTCTTCTGGAATGGTT
2127 ------TGA----------A------A--------A----------

2246 AG....AAGTGAGGGAGTTTGCCCCGGTTCTGTTTGTAGAGTCTCATAGTT
2177 --AACT------A--TC-------A----------CC--.A--------

2292 GGACTTTCTAGCATATATGTGTCCATTTCCTTATGCTGTAAAAGCAAGTC
2225 --------.---TG---T--A----------G----A-------AC-

2342 CTGCAACCAAACTCCCATCAGCCCAATCCCTGATCCCTGATCCCTTCCAC
2274 ---------GCT-T-TG----G-.................-GT----TTG-

2392 CTGCTCTGCTGATGACCCCCCCAGCTTCACTTCTGACTCTTCCCCAGGAA
2308 -----------TT---T--T-TT--TC-TG------GT--C----TG-AG-

2442 GGGAAGGGGGGTCAGAAGAG................AGGGTGAGTCCTCC
2358 -----G-A------------GAATTCTGGAGGATCC---A-AT-----T-

2476 AGAACT....CTTCCTCCAAGGACAGAAGGCTCCTGCCCCCATAGTGGCC
2408 T-----CCTG----T----GT----A-----C---A-TG--CC-...--

2522 TCGAACT...CCTGGCACTACCAAAGGACACTTATCCA.CGAGAGCGCAG
2454 C-A-C--GCC--AT-----C-TCT-------C-T----TACTTTT-A--A

2568 CATCCGACCAGGTTGTCACTGAGAAGATGTTTATTTTGGTCAG.TTGGGT
2504 --C-TAG--------A---...C--T--------G-----T-C----AA

2617 TTTTATGTATTA....TACTTAGTCAAATGTAATGTGGCTTCTGGAATCA
2551 -----CC-G---GGCT---------C---CA-----

2663 TTGTCCAGAGCTGCTTCCCCGTCACCTGGGCGTCATCTGGTCCTGGTAAG
2586 ............-AC---AA--TGGG-ATCCC------.....-T-G---

2713 AGGAGTGCGTGGCCCACCAGGCCCCCCCTGTCACCCATGACAGTTCATTCA
2619 -............--------------------------------AC-T

2763 GGGCCGATGGGGCAGTCGTGGTTGGGAACACAGCATTTCAAGCGTC.ACT
2626 A---T--TA-AT-T-TACTT------TTG-----C--TGG-T-C-GTT-

2812 TTATTTCATTCGGGCCCCACCTGCAGCTCCCTCAAAGAGGCAGTTGCCCA
2676 --C----T-C-T---TTTT-T-------------CCAC--CCAC-A-C

2862 GCCTCTTTCCCT.....TCCAGTTTATTCCAGAGCTGCCAGTGGGG...C
2724 C--CAC--A---GTATG-AG-AC-G---T--A-A----AG----T-GTAA

2904 CTGAGGCTCCTTAGGGTTTTCTCTCTATTTCCCCCTTTCTTCCTCATTCC
2774 ------A-T----------.--TC-C---TTG---CT----C------

2954 CTCGTCTTTCCCAAA....GGCATCACGAGTCAGTCGCCTTTCAGCAGGC
2822 ---A--...-A----TAAG-------A--------T----A--------

3000 AGCCTTGG.CGGTTTATCGCCCTGGCAGGCAGGGGCCCTGCAGCTCTCAT
2869 ---T----TG--G--T-TC------A--C-----A-----......---G

3049 GCTGCCCCTGCCTTGGGGTCAGGTTGACAGGAGGTTGGAGGG.AAAGCCT
2913 ------T--------T----------------------------A------

3098 TAAGCTGCAGGATTCTCACCAGCTGTGTCCGGCCCAGTTTTGGGGTCTGA
2963 ----TCATG-------------------T---T---ACC---AA-G---

3148 CCTCAATTTCAATTTTGTCTGTACTTGAACATTATGAA..GATGGGGGCC
3013 --.......-TT-------.....T........G.A..GT.TG..T..TA

3196 TCTTTCAGTGAATTTGTGAACA..GCAG.AATTGACCGACAGCTTTCCAG
3056 C---AA-C-----A------G-ATC---A--C-----A---------AGA

3243 TACCCATGGGGCTAGGTCATTAAGGCCACATCCACAGTCTCCCCCACCCT
3106 -----.-----------C------T-----.----T---T-----

3293 TGTTCCAGTTGTTAGTTACTACCTCCTCTCCTGACAATACTGTATGTCGT
3151 -----T---------C---------TC-CAGAT-G-T-G-----A--C-

3343 CGAGCTCCCCCCAGGTCTACCCCTCCCGGCCCTGCCTGCTGGTGGGCTTG
3201 -C-A--......AT-A--AT--TGG---AAG-T------T-CT--A-TC--

3393 TCATAGCCAGTGGGATTGCCGGTCTTGACAGCTCAGTGAGCTGGAGATAC
3246 --T--A--------A-C--T-CC-----.GT-TG-------T--AG--CT-

3443 TTGGTCACAGCCAGGCGC...TAGCACAGCTCCCTTCTGTTGATGCTGTA
3295 ----------------T-TAG---T-------T-----C--G-----

3490 TTCCCATATCAAAAGGCACAGGGGACACCCAGAAAACGCCACATCCCCCAA
3345 --T---------------G-T-T-----T----TC-------G

3540 TCCATCAGTGCCAAACTAGCCAACGGCCCCAGCTTCTCAGCTCGCTGGAT
3395 -----------------A----C-T-AT------A--.............

3590 GGCGGAAGCTGCTACTCGTGAGCGCCAGTGCGGGTGCAGACAATCTTCTG
3430 .................-----A--------.....C----

3640 TTGGGTGGCATCATTCCAGGCCCGAAG.CATGAACAGTGCACCTGGGACA
3447 --CA---CT----CAA---A-TA--G-C---------T--G--G-----AC
```

Fig. 3 (1)

09/720215

```
3689 GGGAGCAGCCCCAAATTGTCACCTGCTTCTCTGCCCAGCTTTTCATTGCT
3497 CA-----A---G--T-.....G-----G--T-ATT-------C-G-----

3739 GTGACAGTGATGGCGAAAGAGGGTAATAACCAGACACAAACTGCCAAGTT
3542 C-----A-----A.....A---C-A--C-G-----TTA-A---G-------G---

3789 GGGTGGAGAAAGGAGTTTCTTTAGCTGACAGAATCTCTGAATTTTAAATC
3589 T--CA---------AA---C-----------C-C-----G---------A

3839 ACT..TAGTAAGCGGCTCAAGC....CCAGGAGGGAGCAGAGGGATACGA
3639 GG-TG--A----T-------A-CCAT----AAA----A-A--G-TA--

3883 GCGGAGTCCCCTGCGCGGGACCATCTGGAATTGGTTTAGCCCAAGTGGAG
3689 A-T--....-CAGAT-A----G-CT--T--CA-GC-----A-----

3933 CCTGACAGCCAGAACTCTGTGTCCCCCGTCTAACCACAGCTCCTTTTCCA
3734 T-CAG-T-T-T---------CAG-A-TTC---.--T-----....-C-C-T-

3983 GAGCATTCCAGTCAGGCTCTCTGGGCTGACTGGGCCAGGGGAGGTTACAG
3779 -----------C--------TCA-----.........---A--CA-C----

4033 GTACCAGTTCTTTAAGAAGATCTTTGGGCATATACATTTTTAGCCTGTGT
3821 --G--------C-----CT-----T--.---CAG--CA-----A-A-

4083 CATTGCCCCAAATGGATTCCTGTTTCAAGTTCACACCTGCAGATTCTAGG
3869 -T-----A-.-...----GTA-A----G---A----.--A---------

4133 ACCTGTGTCCTAGACT...........TCAGGGAGTCAGCTGTTTCTAG
3914 G-AGA--A-TG-----CAGAAAAAAAGCC-CT-TG-A-T-TG--A-AGC-

4171 AGTTCCTACCATGGAGTGGGTCTGGAGGA.......CCTGCCCGGTGGGG
3964 -AG-A-A-AA-CT--AG---G--A-G-C-GATGCCG-A-----TCA--CCA

4214 GGGCAGAGCC..CTGCTCCCTCC............GGGTCTTCCTACTCT
4014 -A--CA-----CT-------A---ACATCCTTTCT--C--C-T--T-.--

4250 TCTCTCTG...................CTCTGACGGGATTGTTGATTCT
4063 G-------CTTCAGTGAACCAGCCCCA------A-A-------

4281 CTCCATTTTGGTGTCTTTCTCTTTTAGATATTGTATCAATCTTTAGAAAA
4113 ---------TA------------G--C-A--.........-------

4331 GGCATAGTCTACTTGTTATAAATCGTTAGGATACTGCCTCCCCCAGGGTC
4155 ---T-------A---------------T-C---A---------------

4381 TAAAATTACATATTAGAGGGGAAAAGCTGAACACTGAAGTCAGTTCTCAA
4205 -----A--T--GC--A---------CT-----------AC------G--

4431 CAATTTAGAAGGAAAACCTAGAAAACATTTGGCAGAAAATTACATTTCGA
4255 ----------------T---------AA--A-------T----TAA

4481 TGTTTTTGAATGAATACAAGCAAGCTTTTACAACAGTGCTGATCTAAAAA
4305 -----...-------AG--G-G------GA--A-A--T-------T---

4531 TACTTAGCACTTGGCCTGAGATGCCTGGTGAGCATTACAGGCAAGGGGAA
4351 -------.-TT-A---------G--T--AA----TGAACTGA----T-----

4581                            TCTGAGGTAGCCGACC
4450 GAGATTTGAGGTGTTTTAGCATTGGAAAGCCAC----TTG----T-G--

4598 TGAGGACATGGCTTCTGAACCTGTCTTTTGG.........GAGTGGTATG
4500 CC-.--A-CTA-------C-T----A------AATGGAGGTT------C--

4639 GAAGGTG...............................GAGCG
4549 CC--A--CCAAAGCTGCATGAGACCAGCTCTTGGTTTATCAATTT--A-A

4651 TTCACCAGTGACCTGGAAGGCCCAGCACCACCCTCCTTCCCACTCTTCTC
4599 C-...----A-----A----------A-AGTGT--G--.......-

4701 ATCTTGACAGAGCCTGCCCCAGCGCTGACGTGTCAGGAAAACACCCAGGG
4639 T----A--T-------------A--................---TGCACA

4751 AACTAGGAAGGCACTTCTGCCTGAGGGGCAGCCTGCCTT..GCCCACTCC
4673 --T----G---GT--A--T---ACA-A---T--CT--C-GG----C----

4799 .........TGCTCTGCTCGCCT................CGGA
4723 CATCCTTTG-A-----A-CTA---GACCTTCAGGATCTTGGCACATA--A-

4817 TCAGCTGAG...................CCTTCTGAGCT.........GG
4773 ATG----T-TAGCAAGCACTTTGGCATGC---C--A-A--TACCCCAGA-

4839 CCTCTCACTGCCTCCCCAAGGCCCCCTGCCTGCCCT..............
4823 ------C---------TT---C-AGT-------T-T-CTGGGGAGGTGTTA

4875 .............................GTCAGGAGGCAGAAGGAAGCAGGTG
4873 GAGCCCATAGAATGGAGAGGAGAAA-AA-A-----A---G-C-G------A

4900 TGAGGGCAGTGCAAGGAGGGAGCACAACCCCCAGCTCCCGCTCCGGGCTC
4923 GT-AAAAG-CT-TG-----A-AG---G--T--TAGG-...........

4950 CGACTTGTGCACAGGCAGAGCCCAGACCCTGGAGG....AAATCCTACC
4960 .....-T-------A----GA-T---C----T-T--GAACT--G-G-C-T-

4995 TTTGAATTCAAGAACATTTGGGGAATTTGGAAATCTCTTTGCCCCCAAAC
5005 --G--G--T------------AC--G--C----GA-C--TG-T--TTG-

5045 CCCCATTCTGTCCTACCTTTAATCAGGTCCTGCTCAGCAGTGAGAGCAGA
5055 T--........--TC------T--GG--C-----T----C--A----A-

5095 TGAGGTGAAAAGGCCAAGAGGTTTGGCTCCTGCCCACTGATAGCCCCTCT
5098 --C-C---------A----------------.T---T-

5145 CCCCGCAGTGTTTGTGTGTCAAGTGGCAAAGCTGTTCTTCCTGGTGACCC
5147 ----T----------------------------------T-

5195 TGATTATATCCAGTAACACATAGA...CTGTGCGCATAGGCCTGCTTTGT
5197 ------G---------TT-.--GATT---AT-------T--------A

5242 CTCCTCTATCCTGGGCTTTTGTTTTGCTTTTTAGTTTTGCTTTTAGTTTT
5246 ---T---T-T----------A----T----C-----------------

5292 TCTGTCCCTTTTATTTAACGCACCGACTAGACACACAAAGCAGTTGAATT
5296 C--A-..........--------T-T------A-----------------
```

```
5342 ...........TTTATATATATATATCTGTATATTGCACAATTATAAACTC
5343 TATATATATATA--------------------T----------

5380 ATTTTGCTTGTGGCTCCACACACACAAAAAAAG....ACCTGTTAAAATT
5393 -------------A-G---------C-----AAAA----T------

5426 ATACCTGTTGCTTAATTACAATATTTCTGATAACCATAGCATAGGACAAG
5443 ------------------------------AG---------

5476 GGAAAATA.AAAAAAGAAAAAAAAGAAAAAAAAAACGACAAATCTGTCTGC
5493 ------A-TTT--A---------A-----G---AA----C----------

5525 TGGTCACTTCTTCTGTCCAAGCAGATTCGGTCTTTTCCTCGCTTCTTT
5543 --------AA--------CT---A--.------------G----

5575 CAAGGGCTTTCCTGTGCCAGGTGAAGGAGGCTCCAGGCAGCACCCAGGTT
5592 ---A-A---C-------T-A------A---------T----------

5625 TTGCACTCTTGTTTCTCCCGTGCTTGTGAAAGAGGTCCCAAGGTTCTGGG
5642 ---TG--.-----------TC--.-------G--C------A-------

5675 TGCAG....................GAGCGCTCCCTT
5690 -A---GACAGTTCATTTCAGCATGGGGTCAGGAGACAA----A---------

5692 GACCTGCTGAAGTCCGGAACGTAGTCGGCACAGCCTGGTCGCCTTCCACC
5740 T--A----C--A-A----T----G--A-T----A---....C------

5742 TCT.......GGGAGCTGGAGTCCACTGGGGTGGCCTGACTCCCCCAGTC
5786 ---AGGGATG-------A-CA-G--TG------A--CA------T-....-

5785 CCCCTTCCCGTGACCTGGTCAGGGTGAGCCCATGTGGAGTCAGCCTCGCAG
5833 A----T-.....----C----AA-----TG---AC---A--T--GA--A

5835 GCCT..CCCTGCCAGTAGGG.TCCGAGTGTGTTTCATCCTTCC.CACTCT
5878 --T--TC---A-TT--C----C--A--CA-T-AAA-A--C--AA---T-

5881 GTCGAGCCTGGGGGCTGGAGCGGAGACGGGAGGCCTGGCCTGTCTCGGA
5928 -CT---T--A--A--A--A-A---TA-A--ATT---T--A----CA-G

5930 ACCTGTGAGCTGCACCAGGTAGAACGCCAGGGACCCCAGAATCATGTGCG
5978 GTG---A-----T-----------T----------------C--CA-C-A

5980 TCAGTCCAAGGGGTCCCCTCCAG.GAGTAGTGAAGACTCCAGAAATGTCC
6028 A---C----T-----T---------AA-----------------CA---

6029 CTT..TCTTCTCCCCCCATCCTACGAGTAATTGCATTTGCTTTTGTAATTC
6078 ---TC--------TGC--C-T------C--------------C-

6077 TTAATGAGCAATATCTGCT.....AGAGAGTTTAGCTGTAACAGTTCTTT
6128 ---------T-------AAAAA-A-A-AA------------------

6122 TTG....ATCATCTTTTTTTTAATAATTAGAAACACC..........AAAA
6178 ---CAAA-GG---A--C--A--------A-------CCCCCCCCAA----

6158 AAATCCAGAAACTTGTTCTTCCAAAGCAGAGAGCATTATAATCACCAGGG
6228 --G-------C-----------------------------...---

6208 CCAAAAGCT.TCCCTCCCTGCT...........GTCATTGCTTCTTCT
6275 ------T--G-----A-A-CT--ACCCCATCTCCTCA-G------G------

6244 GAGGCCTGAATCCAAAAGAAAAACAGCCATAGGCCCTTTCAGTGGCCGGG
6325 A-----A----..--C-GC---G-T-TTTG----------GG---A-T---

6294 CTACCCGTGAGCCCTTCGGAGGACCAGGGCTGGGGCAGCCTCTGGGCCCA
6373 ------T--GAG-TC-T---A---..T---------A-------A-A--C

6344 CATCC..GGGGCCAGCTCCGGCGTGTGTTCAGTGTTAGCAGTGGGTCATG
6421 T----TA--C--TT----TA-G-A--AA-----A----T--A-T----CA

6392 ATGCTCTTTCCCACCCAGCCTGGGATAGGGGCAGAGGAGGCGAGGAGGCC
6471 -CAT-A-C----G--G--A--A---G--------A--A--C-AAG--TT

6442 GTTGCCGCTG....ATGTTTGGCCGTGAACAGGTGGGTGTCTGCGTGCGT
6521 --CT--A---CTACT-AC------ACTG------A-----A-CAT--AT--

6488 CCACGTGCGTGTTTTCTGACTGACATGAAATCGACGCCCGAGTTAGCCTC
6571 ---TA---A-------A--G----TG---G--AG-A--A--------T--

6538 ACCCGGTGACCTCTAGCCCTCGCCCGGATGGAGCGGGGCCCACCCGGTTCA
6621 ---T---------A-------T---------A--..-----T-------

6588 GTGTTTCTGGGGAGCTGGACAGTGGAGTGCAAAAGGCTTGCAGAACTTGA
6669 A----------C---------A----------------A----------

6638 AGCCTGCTCCTTCCCTTGCTACCACGGCCTCC.TTTCCGTTTGATTTGTC
6719 -----TT------A-T------G----------T-----A----------

6687 ACTGCTTCAATCAATAACAGCCGCTCCAGAGTCAGTAGTCAATGAATATA
6769 --------G-----------------------TG--------

6737 TGACCAAATATCACCAGGACTGTTACTCAATGTGTGCCGAGCCCCTTGCC.
6819 ------------------------------C-----------T--T

6786 CATGCTGGGCTCCC.GTGTATCTGGACACTGTAACGTGTGCTGTGTTTGC
6869 TG------------T-----C-------------T-----------

6835 TCCCCTTCCCCTTCCTTCTTTGCCCTTTACTTGTCTTTCTGGGGTTTTTTC
6919 --T------T--------C-----------C---------------

6885 TGTTTGGGTTTGGTTTGGTTTTTATTTCTCCTTTTGTGTTCCAAACATGA
6969 -----------------.....-----T------------------

6935 GGTTCTCTCTACTGGTCCTC.TTAACTGTGGTGTTGAGGCTTATATTTGT
7017 ----T--------------T-----------------C-------

6984 GTAATTTTTGGTGGGTGAAAGGAATTTTGCTAAGTAAATCTCTTCTGTGT
7067 ---------------------C----------------------

7034 TTGAACTGAAGTCTGTATTGTAACTATGTTTAAAGTAATTGTTCCAGAGA
7117 -----A--------------------------------

7084 CAAATATTTCTAGACACTTTTTCTTTACAAACAAAAGCATTCGGAGGGAG
7167 -----GC------GT-A-----A---------.-------T-A------
```

Fig. 3 (2)

```
7134 GGGGATGGTGACTGAGATGAGAGGGGAGAGCTGAACAGATGACCCCTGCC
7216 --AAG------A-A---CA-------CA-T-----...T---T-------

7184 CAGATCAGCCAGAAGCCACCCAAAGCAGTGGAGCCCAGGAGTCCCACTCC
7263 ----------------T---A--...---TA---A-T--TT-T---T---

7234 AAGCCAGCAAGCCGAATAGCTGATGTGTTGCCACTTTCCAAGTCACTGCA
7310 ---T--AG-GA-T--..---------T------T----A------AA---

7284 AAACCAGGTTTTGTTCCGCCCAGTGGATTCTTGTTTTGCTTCCCCTCCCC
7358 -------C--...----A----A-------....--------T---T---

7334 CCGAGATTATTACCACCATCCCGTGCTTTTAAGGAAAGGCAAGATTGATG
7401 ...---------T--.........---G-A-................

7384 TTTCCTTGAGGGGAGCCAGGAGGGGATGTGTGTGTGCAGAGCTGAAGAGC
7422 ..-AA-CT---A-T----------A-A-.-A--A-TA-C-....C-----

7434 TGGG......GAGAATGG...GGCTGGGCCCACCCAAGCAGGAGGCTGGG
7465 --T-CTCACT---T-----AAA----T--T-TGAGTTTT-------A--AC

7475 ACGCTCT.GCTGTGGGCACAGGTCAG..GCTAATGT..........TGGC
7515 C-A--G-G-ACA----G-G-A-A---AA-A---AT-AGCCGCTCCC--C-

7512 AGATGCAGCTCTTCCTGGA.CAGGCCAGGTGGTGGGCATT.CTCTCTCCA
7565 TA-GAT-C-----AA-A--TA--T-A---CCA---A----AT-G---A--

7560 AGGTGTGCCCCGTGGGCATTACTGTTTAAGACACTTCCGTCACATCCCAC
7615 ---CA--TTT-AAA-A---G--CAG-C-G--------T----T-CT---T

7610 CCCATCCTCCAGGGCTCAACAC...TGTGACATCTCTATTCCCCACCCTC
7665 GTTGC--C-T--TA-A--GT--TAA-C---T------G----........-

7657 CCCTTCCCAGGGCAATAAAATGACCATGGAGGGGGCTTGCACTCTCTTGG
7708 G--A-G--T---GG----------TAGCA---ACTC---T--.......--CA

7707 CTGTCACCCGATCGCCAGCAAAACTTAGATGTGAGAAAACCCCTTCCCAT
7753 A---G-T-TA--TC--A-----TC-G-GCC-----T-C-A---...GT-

7757 TCCATGGCGAAAACATCTCCTTAGAAAAGCCATTACCCTCATTAGGCATG
7800 --T------A--C-....----C--T-----.TG---TT------GCAG--T

7807 GTTTTGGGCT........CCCAAAACACCTGACAGCCCCTCCCTCCTCTG
7845 ---CCA-C--AATGTAAGAGG--C-G-G-A-TGTT---T-GGAG----..

7849 AGAGGCGGAGAGTGCTGACTGTAGTGACCA.TTGCATGCCGGGTGCAGCA
7893 ..-T-T---C---T--AC-----------G---GC-ATA-TAGTT-TT-

7898 TCTGGAAGAGCTAGGCAGGGTGTCTGCCCCCTCCTGAGTTGAAGTCATGC
7941 G------C-A-C--A---:---ACA----------AA-A-CC-T--TG-A-

7948 TCCCCTGTGCCAGCCCAGAGGCCGAGAGCTATGGACAGCATT...GCCAG
7991 ------C-------.----G--GG-T---A-C--T-G-AT-G-GCA-----

7995 TAACACAGGCCACCCTGTGCAGAAGGGAGCTGGCTCCAGCCTGGAAACCT
8040 -..G---A-------T----.........TT--T-A-----TCAA----TC

8045 GTCTGAGGTTGGGAGAGGTGCACTTGGGGCACAGGGAGAG.GCCGGGACA
8079 A---A---------CA-.---GACAA----G--A--A--AT--A--

8094 CACTTA.....GCTGGAGATGTCTCTAAAAGCCCTGTATCGTATTCACCT
8128 G--C--GCTGG-------GG----TG---------C-G-----------

8139 TCAGTTTTTGTGTTTGGGACAATTACTTTAGAAAATAAGTAGGTCGTTT
8178 -----------C---------------------------------

8189 TAAAAACAAAAATTATTGATTGCTTTTTTGTAGTGTTCAGAA.AAAAGGT
8228 -----------..------------------------A--C-------

8238 TCTTTGTGTATAGCCAAATGACTGAAAGCACTGATATATTTAAAAACAAA
8276 --------------------------------------------

8288 AGGCAATTTATTAAGGAAATTTGTACCATTTCAGTAAACCTGTCTGAATG
8326 --------------------------------------------

8338 TACCTGTATACGTTTCAAAAACACCCCCCCCCCCACTGAATCCCTGTAACC
8376 --------------------....-A----------C----------

8388 TATTTATTATATAAAGAGTTTGCCTTATAAATTTA
8422 -----------------------------------
```

Fig. 3 (3)

dashed line: putative promoter

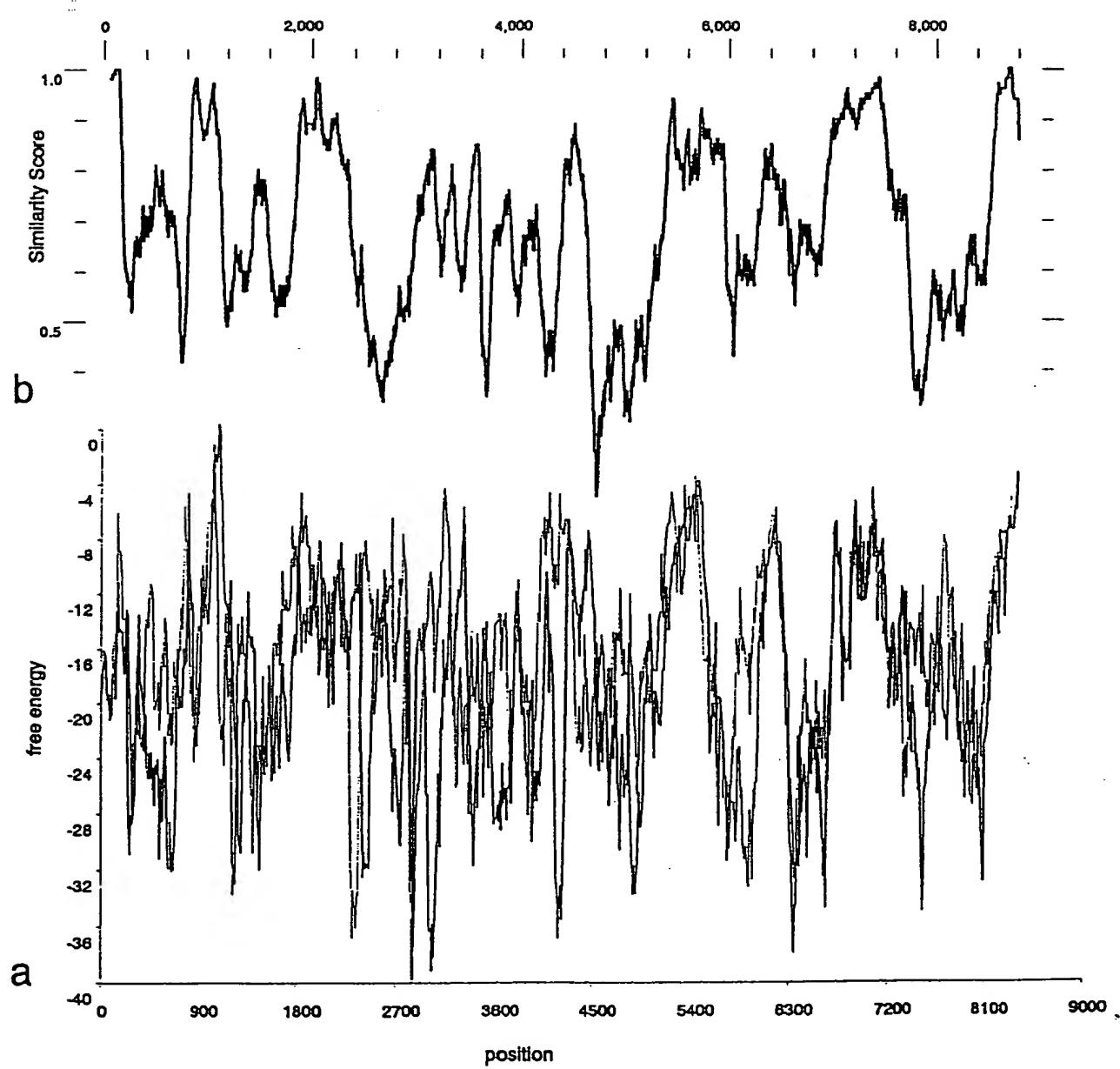full line: sequence-conserved high-energy sequence

Fig. 4

black **similarity 100 window**
blue **hinlex 10** HUMAN

09/720215

```
       1
human  TTGCTGCAGATACTACTGACCAGACAAGCTGTTGACCAGGCACCTCCCCTCCCGCCCAAACCTTT................-...CCCCCATGTGGTCGT
schim  ----------------------------------------------------------------------------------------------------
orang  ------------------------------------------------------------------------..................--.........
makak  ----------------------------------------------------------------------..................--.........
hamst  ----------------------------------------------------..----C..-AATA...-TC...............----A-----C--A-
mouse  ----------------------------------------------TC----CAA-AATATCCT-CC-CTTCCCCCCCCCCCACCCCCG----.G----C---
rat    ----------------------------------------------TC---ACAA-AA-A-CC--C-CCCTCCTCACCCCACCCCTAT----TG----C--A-
kaeng  -----------T------------------------------T-T----TTT-TAGGGTA-A--AGC----GC...................--T----TCATC-C-

       101
human  TAGAGACAGAGCGACAGAGCAGTTGAGAGGACACTCCCGTTTTCGGTGCCATCAGTGCCCCGTCTACA...GCTCCCCCAGCTCCCCCC..ACCTCCCCC
schim  ----------------.........------------.----------------------------------------------......--...---------
orang  -----------------.........-----------.------------------------------------G--...--------------...-----------
makak  -----------------...............-----------------------------------------C...A------------------...-----------
hamst  ------.,.T--------------------A---T------------..T---C-T-------T---TGA---TT-.....-----A----A
mouse  ---......G--------A---A--------A---T--------TGA-----T---C-T-ATA---T---TGA--TTTA-....---A----A
rat    ------..--G--------A--------------A---T-------CTGA-----T---C-T-....---T---TGA--TTTA-...---A----A
kanga  ..............-----G---AGGGA-GGT--AC-G-T--T--CT--A-A--TAA-AGAGTAG~G-TAGTGG-AG-TTA-ATTTT-AGTG--

       201
human  ACTCCCAACCACGTT.GGGACAGGGAGGTGTGAGGCAGGAGAGACAGTT..GGATTCTTTAGAGAAGA...TGGATATGACCAGTGGCTATGGCCTGTGC
schim  ------------------.--------------------------------.--------------...-----------------------T
orang  ------------------.--------------------------------.---------C------...----------------------C-----------
makak  ------------------.--------------------------------.---------G------.....----------C----
hamst  --------T-TGA...------C--------C---------A-A-T-TG-...------CTA..--GG--------A--G----CA.-T
mouse  --------T-TGA...A----T-----------T--------A--TA-G..--C----GG---------CTA-----GT--G------AT--A----CA.CT
rat    --------T-TGA...------T-----------C-------A--TA-ATA---C---GG---------CTA-A--GT-G-A----AT-GC-C--CA.-T
kanga  -T-AA--TT-TA--CCAA-GTCTTA-AT-A-T-T-TT-AG-G-TTTT--...........---CCCT---GG-GCC.G--GGG--GGGG-A--G-.--ATTA-...

       301
human  GATCCCACCCGTGGTGGCTCAAGTCTGGCCCCACACCAGCCCCAATCCAAAACTGGCAAGGACGCTTCACAGGACAGGAAAGTGGCACCTGTCTGCTCC
schim  ----------------------------------------------------------------------------------------------------
orang  ------------------C---------------------------------------------------------------------------------
makak  ------------------C-------A-----------C-----------------------.-----------------------------------T---
hamst  A--TAG-A-T--------A----------------T-CA---.-----------AT-----TAT--TGA-AA---------CA-T--------A--
mouse  A-T---A-T-----A--A--GA------T-----T-A---.-----------A-----TAT--TG--AA-A-A---------------A--
rat    A--T---A-T--A--A-A------------T------A---.-----------A---------AT--TG--AA-A-T----------TT------A--
kanga  AT---TTAGGAAA-A--G-TG--A-A-A-AG--G-G--CTGAGC-GTTGGC---A..------GA-C-TGACTAGGG-CC-G------T-.A---AA---

       401
human  AGCTCTGGCATGGCTAGGAGGGGGGGAGTCCCTTGAACTACTGGGT.GTAGACTGGCCTGAACCACAGGAGAGGATGGCCCAGGGTGAGGTGGCATGGTCC
schim  -----------------------------------------------.----------------------------------------------G-----
orang  -----------------------------A-TC--------------.------------------------------------------------------
makak  ......-------------------A-TT-------------------.------------A-T-------------------------------------
hamst  --------T--A-G------T-A-T...--.------G----A-TAC--------T---G------A---T-----A----A--A-A--AT---CAC-T
mouse  -------A---------AG-T-A-T...--.-A---GA---C-TA-A---A-------G-----A---T---------A---A---T---CA--T
rat    ----------------AGAT-A-T....-T.-A---GT---C-TA-A---------G-A---A--------A--A-A---T---CA--T
kanga  ......CAAGG----CCA--T-A-T.....A--AGGG-GGG--AAGAC-T-A-A-AAGGA-TAG..AA-C-----A--T---CC-A-A-AA-AGC---T.

       501
human  ATTCTCAAGGGACG.TCCTCCAACGGGTGGCGCTAGA.....GGCCATGGAGGCAGTAGGACAAGGTGCAGGCAGGCTGGCCTGGGGTCAGGCCGGGCAG
schim  ---------------.---------------------.....----------------C--------------C-------------
orang  ---------------.-------------AA..-------------------C---------------C-----------------
makak  ----------------.----GTT------A---GA...----------------CA----------CA-------------C-
hamst  GC--A-----C-T-C-T-A-T-GA-AA-AATT---A-GAGG.T-------------C--------CA--A----T----GTG------A---A-A-CT
mouse  G---A-----C-T-C---C-T-GAA-A-AAT---A-GAGG.T-------------C-------A--A---------A-GTG-A-.---A---A--CT
rat    G---A-----C-T-C---C-T-GAA-T-CAT---A-GAAG.T-------------C-------A--A------------A-GTG-A--A---A---CT
kanga  ......--AACC-TAC--A---GGA-T--A-TTG-A-GAGGCCC-T-------A-TCCCC-ACCACCAA-A---.....AT-T--A-C-GCA--T.--TT

       601
human  AGCACAGCGGGGTGAGAGGGATTCCTAATCACTCAGAGCAGTCTGTGACT..............TAGTGGACAGGGGAGGGGGCAAAGGGGGAGGAGAAG
schim  ---------------------------------------------------.----------------------------------------------
orang  G---------------------------------G-------GGTAGTTAGGGACTC----------------.-G-G----C--------
makak  -A-----------------------CGT--.-------------CGGTAGTTAGGGACT-----------------A-----.----------------
hamst  --TG-CA-A-AACA----A-CAAT--G.TG---...--T---TA-G--------T-----TG--AC-A---.GC-T------A--A-T-
mouse  --TG-CA-A-AACA----A-CAGT--..G---...--T-A-TAAGA.......----------------A--A-G-
rat    --TG-CA-A-A-CA----A-CAGT--C.TG---...--T---TAAG-.....----------------A--A-GA
kanga  CATTT--T--ACCTT-T-TATA--TGGGTGTG-ATGCAC-TAGATA---.............A-TGA--A-GA

       701
human  AAAATGTTCTTCCAGTTACTTTCCAATTCT...CCTTTAGGGACAGCTTAGAATTATTTGCACTATTGAGTCTTCAT....GTTCCCACTTCAAAACAAA
schim  -----------------------------...-.----------------------------------------....-------------------
orang  -----------------------------...-.----------------------------------------....-------------------
makak  -------------------------ACT...-.----------------------------------------....-------------------
hamst  -----A---------A---------G----...-.-------------------------------------....-----A--------
mouse  -----A---------A---------G----...-.-------------------------------------....-----------------
rat    -----A---------A--G------G----...-.-------------------------------------....-----------------
kanga  ---C-------------CT------G-G--ACTAGC----------G------C-G-C-----A----T--GA----TCCCA----------------

       801
human  CAGATGC.....TCTGAGAGCAAACTGGCTTGAATTGGTGACATTTAGTCCCTCAAGCCACCAGATG......TGACAGTGTTGAGAACTACCTGGATTT
schim  -------.....----------------------------------------.....--.------------------------
orang  -.-----.....--------------------------------.....--.----------------A-
makak  -------.....-----------------------G-------------G------------...--TG-------------------C-
hamst  -------.....-----A---------------A-------C--T---A----------A---...--G-------T----------
mouse  -------.....-----A---------------A-------C-..----A--------CA..---G-------C----------T--C-
rat    T------....-----A---------------A-------C-..----A--------CAA.....--G-------C----------T--A-
kanga  -T--CTGAGATG-TCA-C--A----------C---G--A----GCCC-TG-CACTTA-TTA---CACTGGTG---TG-G---TT-----C--AT----

       901
human  GTATATATACCTG                        Fig. 5
schim  -------------
orang  -------------
makak  -------------
hamst  -------CCTG..
mouse  -------......
rat    -------CCTG..
kanga  ---GG---CTG..
```

Partial sequence of the non-coding RNA gene from hamster

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCCCCCCA
 51  ATACTCCCCC AATGTGCTCA TTAGAGATAG CAGTTGAGAG GACACTCCCA
101  TTTTTGGTGC CCTGTCCATA GCTTCCCTGA CTCTTCCACC ACCCCAACTC
151  CCAATCTGAG GGACCGGGAG GTGCGAGGCA GGAAAAATAT TGGATTCTTT
201  AGAGAAGACT AGAGGTGACC AGTGACTGTG GCCCAGTAAT TAGAACTGTG
251  GTGGCACAAG TCTGGCCCCA CATCCACCCA ATCCAAAACT GATAAGGATA
301  TTTTGAAAAA CAGGAAAGCA GTACCTGTCT GATCCAGCTC TGGTATAGGT
351  AGGAGTGAGT CCTGAACTGC TGGATTACAG ACTGGCTTGA GCCACAGAAG
401  ATGATGGACC AGAGTAAAGT ATCATCACCT GCTCACAAGG CATGCTTCAC
451  TAGAGAATAA TTCTAAAGAG GTGCCATGGA GGCAGCAGGA CAAGGCACAA
501  GCAGTCTGGG TGGGGGTCAA GCCAGACCTA GTGCCACAGA ACAAGAGAGC
551  AATCTGTGAC TAGTAGTTAG GGACTTTGTG GATGGGACAA GGGGCATGGG
601  GGAAGAAATG AAAATATTCT TCCAATTACT TTCCAGTTCT CCTTTAGGGA
651  CAGCTTAGAA TTATTTGCAC TATTGAGTCT TCATGTTCCC ACTTAAAAAC
701  AAACAGATGC TCTGAAAGCA AACTGGCTTG AAATGGTGAC ACTTTGTCCC
751  ACAAGCCACC AAATGTGGCA GTGTTTAGAA CTACCTGGAT CTGTATATAC
801  CTG
```

Fig. 5a

Partial sequence of the non-coding RNA gene from kangaroo

```
  1  TTGCTGCATA TACTACTGAC CAGACAAGCT GTTTATCAGG CTTTTTAGGG
 51  TACACCAGCA CCTGCCCTCC ATTCATCCCT GTTGGGAGAG GGATGGTGTA
101  CTGGTTGTCA CTAGAGACCT AACAGAGTAG GGTTAGTGGG AGCTTACATT
151  TTCAGTGCCA TTAACATTCT AGTCCAAGGT CTTAAATTAT TATGTTGAGG
201  GGTTTTTTTT CCCCTGAGGG GGCCGGGGGG TGGGGGGAGG GTTGATTAGA
251  TTCCTTAGGA AAGAGGGTTG AGACAGACAG CAGAGCACTG AGCAGTTGGC
301  ACTAAAGGAG ACCTTGACTA GGGGCCAGGT GGCATCATCT AATCCCAAGG
351  GGCTCCAAGT GAGTATTAGG GTGGGGGAAG ACATTATAGA AGGAATAGAA
401  ACAGGATAGC TCAGCCTAAA GAAGAGCGGT TAAAACCCTA CCCACCAGGA
451  GTTGACTTGA AAGAGGCCCC TATGGAGGAA TCCCCAACCA CCAAAAGCAA
501  TCTTGAGCTG CAGCTGCTTC ATTTAGTGGA CCTTGTGTAT ATCTGGGTGT
551  GTATGCACAT AGATAGACAG TGAGAAAGAA AACTGTTCTT CCAGTTCTTT
601  TCCAGTGCTA CTAGCTTAGG GACAGGTTAG AACTGTCTGC ACAATTGTGT
651  GATCATTCCC ATTCCCACTT CAAAACAAAC TGACTGAGAT GTTCAACAGA
701  AAACTGGCTT CAATGGGTAA CATGCCCTTG CCACTTACTT AAGACACTGG
751  TGTGATGGGG TTTTGAACTC CCTATATTTG TAGGTATCTG
```

Fig. 5b

Partial sequence of the non-coding RNA gene from makaka

```
  1    TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT
 51    CCCGCCCAAA CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCAGTTGAGA
101    GGACACTCCC GTTTTCGGTG CCATCAGTGC CCCGTCTACC ACTCCCCCAG
151    CTCCCCCCAC CTCCCCCACT CCCAACCACG TTGGGACAGG GAGGTGTGAG
201    GCAGGAGAGA CAGTTGGATT CTTTAGAGAT GGATGTGACC AGTGGCTATG
251    GCCCGTGCGA TCCCACCCGT GGCGGCTCAA ATCTGGCCCC ACCCCAGCCC
301    CAATCCAAAA CTGGCAAGGA CGCTTCACAG GACAGGAAAG TGGCACCTGT
351    CTGTTCCGGC ATGGCTAGGA GGGAGTTGTC CCTTGAACTA CTGGGTGTAG
401    ACTGGCCTAA ATCACAGGAG AGGATGGCCC AGGGTGAGGT GGCATGGTCC
451    ATTCTCAAGG GACGTCCTCC AGTTGGTGGC ACTAGAGAGG CCATGGAGGC
501    AGTAGGACAA GGCACAGGCA GGCTGGCCCA GGGTCAGGCC GGGCCGAACA
551    CAGCGGGGTG AGAGGGATTC CTCGTCTCAG AGCAGTCTGT GACCGGTAGT
601    TAGGGACTTA GTGGACAGGG AAGGGGCAAA GGGGGAGGAG AAGAAAATGT
651    TCTTCCAGTT ACTTTCCAAT TCTACTCCTT TAGGGACAGC TTAGAATTAT
701    TTGCACTATT GAGTCTTCAT GTTCCCACTT CAAAACAAAC AGATGCTCTG
751    AGAGCAAACT GGCTTGAATT GGTGACGTTT AGTCCCTCAG GCCACCAGAT
801    GTGATGGTGT TGAGAACTAC CTGGATATGT ATATATACCT G
```

Fig. 5c

Partial sequence of the non-coding RNA gene from orangutan

```
  1   TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT
 51   CCCGCCCAAA CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCAGTTGAGA
101   GGACACTCCC GTTTTCGGTG CCATCAGTGC CCCGTCTGCA GCTCCCCCAG
151   CTCCCCCCAC CTCCCCCACT CCCAACCACG TTGGGACAGG GAGGTGTGAG
201   GCAGGAGAGA CAGTTGGATT CTTTCGAGAA GATGGATATG ACCAGTGGCC
251   ATGGCCTGTG CGATCCCACC CGTGGCGGCT CAAGTCTGGC CCCACACCAG
301   CCCCAATCCA AAACTGGCAA GGACGCTTCA CAGGACAGGA AAGTGGCACC
351   TGTCTGCTCC AGCTCTGGCA TGGCTAGGAG GGAGTCGTCC CTTGAACTAC
401   TGGGTGTAGA CTGGCCTGAA CCACAGGAGA GGATGGCCCA GGGTGAGGTG
451   GCATGGTCCA TTCTCAAGGG ACGTCCTCCA ACGGGTGGCG CTAGAAAGGC
501   CATGGAGGCA GTAGGACAAG GCGCAGGCAG GCTGGCCCGG GGTCAGGCCG
551   GGCAGGGCAC AGCGGGGTGA GAGGGATTCC TAATCACTCA GAGCAGTGTG
601   TGACTGGTAG TTAGGGACTC AGTGGACAGG GGAGGGGCGA GGGGGCAGGA
651   GAAGAAAATG TTCTTCCAGT TACTTTCCAA TTCTCCTTTA GGGACAGCTT
701   AGAATTATTT GCACTATTGA GTCTTCATGT TCCCACTTCA AAACAAACGA
751   TGCTCTGAGA GCAAACTGGC TTGAATTGGT GACATTTAGT CCCTCAAGCC
801   ACCAGATGTG AGTGTTGAGA ACTACCTGGA TTTGTATATA TACCTG
```

Fig. 5d

Partial sequence of the non-coding RNA gene from rat

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACTCCCCAC

 51  AACAACAACC CCCTCCCTCC TCACCCCACC CCTATCCCCT GTGTGCTCAT

101  TAGAGAGGGC AATTGAGAGG ACACTCCCAT TTTTGGTGCC ACTGATGCCC

151  TGTCCATAGC TTCCCTGACT TTTACACCAC CCCAACTCCC AATCTGAGGG

201  ACTGGGAGGT GTGACGCAGG AGAAACTATA TAGGACTCTT GGGAGAAGAC

251  TATAGAGTTG GCAAGTGATT GCGCCCCAGT AATTCCAACT GTGGTAGCAC

301  AAGTCTGGCT CCACACCAAC CCAATCCAAA ACTGACAAGG ACATTTTGCA

351  AAAAATGAAA GTGGCATTTG TCTGATCCAG CTCTGGCATG GCTAGAGATG

401  AGTCTTAAAC TGTTGGCTTA TAAACTGGCC TGAGCAACAG AAGAGGATGG

451  CCCAGAGTAA AGTGTCATCA TCTGTTCACA AGGCATGCTC CCCTAGAAGT

501  TCATGCTAAA GAAGTGCCAT GGAGGCAGCA GGACAAAGTA CAGGCTAGGT

551  GGAGTCAAGC CAGGCCTAGT GCCACAGAGC AAGAGAGCAG TCTCTGACTA

601  GTAGTTAAGG GGGAAGAAAG AAAAATATTC TTCCAATTGC TTTCCAGTTC

651  TCCTTTAGGG ACAGCTTAGA ATTATTTGCA CTATTGAGTC TTCATGTTCC

701  CACTTCAAAA CAAATAGATG CTCTGAAAGC AAACTGGCTT GAAATGGTGA

751  CACTGTCCCA CAAGCCACCA GACAATGGCA GTGTTCAGAA CTACCTGTAT

801  ATGTATATAC CTG
```

Fig. 5e

09/720215

Partial sequence of the non-coding RNA gene from chimpanzee

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT
 51  CCCGCCCAAA CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCGACAGAGC
101  AGTTGAGAGG ACACTCCCGT TTTCGGTGCC ATCAGTGCCC CGTCTACAGC
151  TCCCCCAGCT CCCCCCACCT CCCCCACTCC CAACCACGTT GGGACAGGGA
201  GGTGTGAGGC AGGAGAGACA GTTGGATTCT TTAGAGAAGA TGGATATGAC
251  CAGTGGCTAT GGCCTGTGTG ATCCCACCCG TGGTGGCTCA AGTCTGGCCC
301  CACACCAGCC CCAATCCAAA ACTGGCAAGG ACGCTTCACA GGACAGGAAA
351  GTGGCACCTG TCTGCTCCAG CTCTGGCATG GCTAGGAGGG GGGAGTCCCT
401  TGAACTACTG GGTGTAGACT GGCCTGAACC ACAGGAGAGG ATGGCCCAGG
451  GTGAGGTGGC GTGGTCCATT CTCAAGGGAC GTCCTCCAAC GGGTGGCGCT
501  AGAGGCCATG GAGGCAGTAG GACAAGGCGC AGGCAGGCTG GCCCGGGGTC
551  AGGCCGGGCA GAGCACAGCG GGGTGAGAGG GATTCCTAAT CACTCAGAGC
601  AGTCTGTGAC TTAGTGGACA GGGGAGGGGG CAAAGGGGGA GGAGAAGAAA
651  ATGTTCTTCC AGTTACTTTC CAATTCTCCT TTAGGGACAG CTTAGAATTA
701  TTTGCACTAT TGAGTCTTCA TGTTCCCACT TCAAAACAAA CAGATGCTCT
751  GAGAGCAAAC TGGCTTGAAT TGGTGACATT TAGTCCCTCA AGCCACCAGA
801  TGTGACAGTG TTGAGAACTA CCTGGATTTG TATATATACC TG
```

Fig. 5f